Failure Diagnosis of Discrete-Time Stochastic Systems subject to Temporal Logic Correctness Requirements

Jun Chen, *Student Member, IEEE* and Ratnesh Kumar, *Fellow, IEEE*
Dept. of Elec. & Comp. Eng., Iowa State University, Ames IA 50011, USA
(e-mail: junchen,rkumar@iastate.edu)

***Abstract*— This paper studies the failure diagnosis of discrete-time stochastic systems with linear-time temporal logic (LTL) as correctness requirement—A fault is a violation of the LTL specification. The detection problem is first reduced to stochastic reachability estimation problem for an input-output stochastic hybrid automaton (I/O-SHA) introduced in this paper, based on which the *likelihood of no-fault* is recursively computed for issuing a detection decision. The performance of the detection scheme is measured in terms of false alarm (FA) and missed detection (MD) rates, and the condition for the existence of a detector to achieve any desired rates of FA and MD is captured in form of *Stochastic-Diagnosability*. The proposed method of fault detection is illustrated by a practical example.**

I. Introduction

The problem of fault detection has been recently studied in the setting of stochastic systems [1], [2], [3], [4], [5], [6]. In this paper we study fault detection of certain physical systems modeled as stochastic difference equations, where a fault is a violation of certain correctness requirements expressed as linear-time temporal logic (LTL) formulas.

We first introduce the notion of input-output stochastic hybrid automaton (I/O-SHA), extending the logical input-output hybrid automaton (I/O-HA) introduced in [7], and propose an algorithm that refines the stochastic systems against the LTL formula to yield an I/O-SHA. The likelihood of fault versus no-fault (requirement-violation versus non-violation) is recursively computed and is used as a statistic for issuing fault detection decisions: Whenever the likelihood of no-fault falls below a suitable threshold, i.e., the likelihood of no-fault is "low", a fault decision is issued, and otherwise the detector remains silent. The performance of this detection scheme is determined by introducing and computing its false alarm (FA) and missed detection (MD) rates. In order to identify the class of systems for which detection with any desired accuracy is feasible, we introduce the notion of *Stochastic-Diagnosability* as the corresponding necessary and sufficient condition. The proposed diagnosis framework is implemented for a benchmark room heating problem [6], [8] to demonstrate the validity and applicability of the results.

II. Preliminaries

In this paper, we study fault detection of physical systems subject to disturbance and noise, modeled by stochastic difference equations (1)-(3):

$$x_{k+1} = f(x_k, u_k, v_k), \quad (1)$$
$$r_k = g(x_k, u_k), \quad (2)$$
$$y_k = h(x_k, u_k, w_k), \quad (3)$$

where u, x, r, y, v, w represent, respectively, the input, state, requirement (unobserved), output (observed), disturbance and noise variables, and k is the time-index. Note the requirement variable, being user-defined, is independent of disturbance or noise. The properties of the nonfault system behaviors are described by using a LTL formula over the requirement variables. In the following we present a brief discription of LTL; a more thorough introduction can be found in [9].

Let $M_d = (L_d, \delta, AP, label)$ be a state transition graph, where L_d is the set of states, $\delta : L_d \to 2^{L_d}$ is a total transition relation, i.e., $\forall l \in L_d, \delta(l) \neq \emptyset$, AP is a finite set of atomic proposition symbols, and $label : L_d \to 2^{AP}$ is a function that labels each state with the set of atomic propositions true at that state. A sequence of states $\pi = (l_0(\pi), l_1(\pi), \dots)$ is a *state-trace* in M_d if $l_{i+1}(\pi) \in \delta(l_i(\pi))$ for every $i \in \{0, 1, \dots\}$. $\pi^k = (l_k(\pi), l_{k+1}(\pi), \dots)$, where $k \in \mathbb{N}$, is used to denote the suffix of π starting from index k. A *proposition-trace* over an atomic proposition set AP is defined as a sequence of set of atomic propositions, $\pi_p = (label_0, label_1, \dots)$ such that $label_i \subseteq AP, \forall i \in \{0, 1, \dots\}$. A proposition-trace $\pi_p = (label_0, label_1, \dots)$ over AP is said to be *contained* in M_d if there exists a state-trace $\pi = (l_0, l_1, \dots)$ in M_d such that $label_i = label(l_i), \forall i \in \{0, 1, \dots\}$, in which case π_p is said to be associated with π.

LTL temporal logic is a formalism for describing properties of sequences of states. Such properties are expressed using *temporal operators* of the temporal logic which include: X ("next time"), U ("until"), F ("eventually" or "in the future"), G ("always" or "globally") and B ("before"). We have the following relations among the above operators, where ϕ denotes a temporal logic formula: $F\phi \equiv trueU\phi$, $G\phi \equiv \neg F \neg \phi$, and $\phi B g \equiv \neg(\neg \phi U g)$. So we can use X and U to express all the other temporal operators.

The semantics of LTL can be defined with respect to the *infinite* state-traces in a state transition graph $M_d = (L_d, \delta, AP, label)$. For a LTL formula ϕ, we use the notation $< M_d, \pi > \models f$ (resp., $< M_d, \pi > \not\models f$) to denote that f

The research was supported in part by the National Science Foundation under the grants, NSF-ECCS-0801763, NSF-ECCS-0926029, and NSF-CCF-1331390.

978-1-4799-3106-4/14/$31.00 ©2014 IEEE

holds (resp., does not hold) along the infinite state-trace π in M_d. The detailed definition of the relation $\models$ is omitted here. The semantics of LTL formulas can also be expressed over infinite length proposition-traces without referring to any specific state transition graph.

Given a LTL formula ϕ, denote S_ϕ as the set of all infinitely long proposition-traces over AP satisfying ϕ. Then we can obtain a generalized nondeterministic Büchi automaton T_ϕ ([9]) that accepts S_ϕ. To construct T_ϕ, we first put ϕ into *negation normal form*, in which negation is only applied at the atomic level. Then we rewrite each subformula of the form Fg as $TrueUg$. Let $|\phi|$ be the number of subformulas of the form $\lambda U \mu$. Then the generalized nondeterministic Büchi automaton has $|\phi|$ sets of accepting states and is of the form: $T_\phi = (L_\phi, 2^{AP}, \delta_\phi, l_0^\phi, \mathcal{L}_\phi)$, where $\mathcal{L}_\phi \subseteq 2^{L_\phi}$ is the generalized Büchi acceptance condition, such that for each subformula of the form $\lambda U \mu$ in ϕ, there exists a $\mathcal{L} \in \mathcal{L}_\phi$ which is used to capture the fulfillment of $\lambda U \mu$.

While every LTL formula can be characterized as the ω-language accepted by a nondeterministic Büchi automaton, only certain fragments of LTL can be captured by a *deterministic* Büchi automaton. In this paper we only consider *prediagnosable* LTL formulas (see Definition 1 in next section) that can be accepted by deterministic Büchi automata.

III. Fault Diagnosis Problem Formulation

Suppose the dynamics of a physical system G under diagnosis can be described by the stochastic difference equations (1)-(3), where recall that u, x, r, y, v, w represent, respectively, the input, state, requirement (unobserved), output (observed), disturbance and noise variables, and k is the time-index. The initial state x_0, the *disturbance* v_k as well as the *noise* w_k are all assumed mutually i.i.d. with known distributions. Note the requirement variable, which specifies a required value for each input-state pair through the function g, is used to capture a user-defined specification that, at each step, depends on system state and input, and being a user-defined requirement, it is not corrupted by noise. We assume that the properties of the required system behaviors can be described by using a LTL formula ϕ involving *predicates* defined over the requirement variables r_k, $k \in N$. Then the predicates, appearing in the LTL specification, and their boolean combinations act as atomic propositions guarding the transitions in the Büchi automaton. The set of all infinitely long feasible sequences of aforementioned predicates is denoted a A_G.

Since detection of requirement-violation must occur based on a finite history of input/output observations, it is natural to assume that every infinite run of a system, that violates the given LTL formula, possesses a finite prefix, called an *indicator*, such that all its infinite extensions that are feasible in the system also violate the LTL formula. This property was captured under the name of *prediagnosability* in [10], [11], and is a *necessary* condition for any detector's ability to detect the violation of the specified LTL formula based on finite-length observations. So, without loss of generality, we assume that the prediagnosability holds. Next we provide a formal definition of indicator and also of prediagnosability.

Definition 1: Given a system G and a LTL formula ϕ, a finite sequence of requirement variables is said to be an *indicator* if all of its infinite extensions in G violate ϕ. Denote the set of all indicators as $I_\phi(G)$. G is said to be *prediagnosable* with respect to ϕ if each infinite sequence of requirement variables violating ϕ possesses a finite prefix that is an indicator.

Remark 1: Note that a system is inherently prediagnosable if the LTL formula ϕ is a safety one [9], i.e., it only requires that some "bad" things must never occur. However, when the correctness requirement is a more general one, the system may not be prediagnosable (See Example 1), and in this case, the violation of ϕ can not be detected even if the system is perfectly observable, i.e., $y_k = r_k$ for all $k \in \mathbb{N}$. For this reason, we assume without loss of generality that the system is prediagnosable with respect to the LTL formula. ■

As established in [10, Theorem 1], the prediagnosability of system G with respect to a LTL formula ϕ, is equivalent to the existence of a deterministic Büchi automaton accepting $S_\phi \cap A_G$, which can also be characterized as the *limits* of the finite prefixes accepted by the same model treated as a standard finite state automaton. Then we can augment the Büchi automaton, by adding an absorbing state called F reaching which indicates the execution of an indicator, to yield an augmented deterministic requirement model, denoted R. (Note the augmentation requires adding the "missing" transitions from each state to the newly added fault state F, guarded by the complement of the existing transitions of the state.)

Example 1: Consider a system G with dynamics: $x_{k+1} = x_k + v_k$ and requirement variable $r_k = 2x_k - 1$, where v_k is i.i.d. Gaussian random variable. Suppose the LTL formula is given as $\phi = GF(r < 0)$, i.e., it is always (G) possible that in future (F), the requirement variable becomes negative. Then it can be verified that for any infinite sequence $(r_0, r_1, \dots, r_m, \dots)$ with $r_i \geq 0, \forall i \geq m$ (i.e., a sequence violating ϕ), any of its prefix has certain infinite extension in which $(r_k < 0)$ is satisfied for infinitely many k (i.e., a sequence satisfying ϕ). Therefore G is not prediagnosable with respect to ϕ. In this case even with perfect observation $y_k = r_k$, the violation of ϕ cannot be detected. Now consider the disturbance to be $v_k = sign(x_k)v'_k$, where v'_k is a positive-valued random variable, i.e., the noise v_k is dependent on the state variable x_k and is negative (resp., positive) if x_k is negative (resp., positive). Consider again the LTL formula $\phi = GF(r < 0)$. Then in this case, for every infinite sequence $(r_0, r_1, \dots, r_m, \dots)$ with $r_i \geq 0, \forall i \geq m$ (i.e., a sequence violating ϕ), there exists a finite prefix $(r_0, \dots, r_k)$ with $r_k \geq 0$ (so that $x_k = (r_k + 1)/2 \geq 0.5$) whose all infinite extensions also violate ϕ. Then G is prediagnosable with respect to $GF(r < 0)$. ■

IV. Approach to Detection Problem

Consider the detection structure of Fig. 1. At any given time, the true state of the requirement model R is not avail-

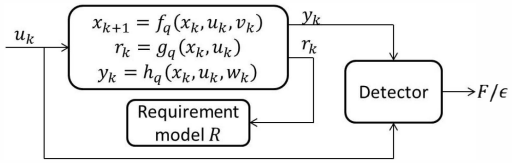

Fig. 1. The detection structure.

able to the detector and must be estimated from the observed history of inputs and outputs. We transform this problem of estimating requirements violation to fault-location rechability estimation in an input-output stochastic hybrid automaton (I/O-SHA) model that captures the behaviors of both G and R in an unified manner.

We first introduce the notion of an I/O-SHA, extending that of a logical input-output hybrid automaton (I/O-HA) given in [7].

A. Input-Output Stochastic Hybrid Automaton

Definition 2: An input-output stochastic hybrid automaton (I/O-SHA) is a 10-tuple $P = (L, D, U, Y, \Sigma, \Delta, \ell_0, d_0, L_m, E)$, where

- L is the set of locations (symbolic states), and each $l \in L$ is a 3-tuple $l = (G_l, f_l, h_l)$, where
  - $G_l : D \times U \to [0,1]$ is the location invariant probability satisfying (4) below,
  - $f_l : D \times U \times D \to [0,1]$ assigns for each $(d,u) \in D \times U$ a probability density function $f_l(\cdot|d,u)$ on the data space D, and
  - $h_l : D \times U \times Y \to [0,1]$ assigns for each $(d,u) \in D \times U$ a probability density function $h_l(\cdot|d,u)$ on the output space Y.
- $D = D_1 \times \cdots \times D_n \subseteq \mathbb{R}^n$ is the set of data (numerical states), and hence the hybrid state space of P is given by $L \times D$,
- $U = U_1 \times \cdots \times U_m \subseteq \mathbb{R}^m$ is the set of numerical inputs,
- $Y = Y_1 \times \cdots \times Y_p \subseteq \mathbb{R}^p$ is the set of numerical outputs,
- Σ is the set of symbolic inputs,
- Δ is the set of symbolic outputs,
- $\ell_0 : L \to [0,1]$ is the initial probability distribution for the locations,
- $d_0 : D \to [0,1]$ is the initial probability distribution for the data values,
- $L_m \subseteq L$ is the set of final locations,
- E is the set of edges (transitions), and each $e \in E$ is a 7-tuple $e = (o_e, t_e, \sigma_e, \delta_e, G_e, f_e, h_e)$, where
  - $o_e \in L$ is the original location,
  - $t_e \in L$ is the terminal location,
  - $\sigma_e \in \Sigma \cup \{\epsilon\}$ is the symbolic input,
  - $\delta_e \in \Delta \cup \{\epsilon\}$ is the symbolic output,
  - $G_e : D \times U \to [0,1]$ is the guard probability satisfying (4) below,
  - $f_e : D \times U \times D \to [0,1]$ assigns for each $(d,u) \in D \times U$ a probability density function $f_e(\cdot|d,u)$ on the data space D,
  - $h_e : D \times U \times Y \to [0,1]$ assigns for each $(d,u) \in D \times U$ a probability density function $h_e(\cdot|d,u)$ on the output space Y.

Remark 2: In Definition 2, G_l and G_e, where $l \in L, e \in E$, capture the probabilities that an I/O-SHA stays in the current location l or executes a transition e, and so it satisfies the following stochasticity constraint:

$$\forall (d,u) \in D \times U, \sigma \in \Sigma \cup \{\epsilon\}, \quad G_l(d,u) + \sum_{e \in E : \sigma_e = \sigma} G_e(d,u) \leq 1. \tag{4}$$

Note that in certain special setting, the range space of G_l and G_e can simply be the binary set $\{0,1\}$ [7], i.e., given any (d,u), an I/O-SHA will either stay at current location, or execute one transition, with probability 1. Then the guard/invariant can be equivalently written as logical predicates, $\overline{G}_l := \{(d,u) : G_l(d,u) = 1\} \subseteq D \times U$ and $\overline{G}_e := \{(d,u) : G_e(d,u) = 1\} \subseteq D \times U$. Since in this paper, we consider refinement of physical systems against their logical LTL formula, only logical guards/invariants are needed in the refined I/O-SHA models. ■

An I/O-SHA P starts from an initial distribution ℓ_0 over L and an initial distribution d_0 over D. At each time step, given a current location l, current data value d and input value u, upon the arrival of a symbolic input $\sigma \in \Sigma \cup \{\epsilon\}$, P evolves either within the current location with probability $G_l(d,u)$ or executes an outgoing edge e such that $\sigma_e = \sigma$ with probability $G_e(d,u)$. In the former case, it updates the data variable d according to the distribution $f_l(\cdot|d,u)$, and the output variable y is assigned a value according to the distribution $h_l(\cdot|d,u)$. In the latter case, the distributions $f_e(\cdot|d,u)$ and $h_e(\cdot|d,u)$ are used for updating d and y, and a symbolic output δ_e is emitted.

Remark 3: In [12], the authors proposed discrete time stochastic hybrid systems (DTSHS), which includes hybrid state/control space. The I/O-SHA model introduced here is more general than the DTSHS model: state variables of a DTSHS are fully observed, whereas the data of an I/O-SHA is only partially and unreliably observed. ■

Next we present the refinement of a system against its LTL formula. Given a physical system G with dynamics described by (1)-(3) and the requirement model R, the refinement is modeled by an I/O-SHA G^R, where

- L is given by the state space of R, $l_0 = \delta(l_0^\phi)$ where δ is the Diract delta function, d_0 is the initial distribution of x_0, and $L_m = \{F\}$,
- D, U, Y are given by the state/input/output space of G, respectively, and $\Sigma = \Delta = \emptyset$,
- the discrete transition structure of G^R is preserved from that of R,
- for each location $l \in L$,
  - location invariant $\overline{G}_l$ is given by $\overline{G}_l = \{(d,u) : g(d,u)$ violates the predicates over each outgoing transition from l in $R\}$,
  - probability density functions $f_l(\cdot|d,u)$ and $h_l(\cdot|d,u)$ for data updates and output assignments

are determined by the distributions of v_k and w_k, together with the functions f and h of G,

- for each $e = (l, l', \sigma_e, \delta_e, \overline{G}_e, f_e, h_e)$, e is a transition of G^R (i.e., $e \in E$), if and only if,
  - there exists a transition of R from l to l', and
  - $\overline{G}_e = \{(d, u) : g(d, u)$ satisfies the predicates over the above transition of $R\}$, and
  - $\sigma_e = \delta_e = \epsilon$, $f_e(d_r|d, u) = \delta(d_r - d)$, and $h_e(\cdot|d, u)$ is the identity function that keeps the output values unchanged on discrete transitions.

Remark 4: The refinement G^R captures the behaviors of both G and R in an unified manner such that any system run associated with an indicator, transitions G^R to the fault-location $L_m = \{F\}$. ■

B. Detection Statistics and Detection Scheme

Denote the history of observed inputs/outputs up to a time k as $u^k = (u_0, \ldots, u_k)$, $y^k = (y_0, \ldots, y_k)$ and let $z^k = (y^k, u^k)$. Define $\pi_k(\cdot|z^k) := Pr(l_k = l|z^k)$ as the conditional probability distribution over the discrete locations given the observations until time k, $p_{k|k}(d_k|z^k, l_{k-1}) := p_{d_k|z^k,l_{k-1}}(d_k|z^k, l_{k-1})$ as the probability distribution function over continuous variables at time k, given z^k and l_{k-1}, and $p_{k+1|k}(d_{k+1}|z^k, l_k) := p_{d_{k+1}|z^k,l_k}(d_{k+1}|z^k, l_k)$ as the probability distribution function over continuous variables at time $k+1$, given z^k and l_k. Note that $p_{k|k}(d_k|z^k, l_{k-1})$ (resp., $p_{k+1|k}(d_{k+1}|z^k, l_k)$) can be interpreted as the posterior (resp., prior) distribution of the data d_k (resp., d_{k+1}) given the input/output up to time k. The following equations (5)-(9) initialize and recursively update the state distributions π_k, $p_{k|k}$ and $p_{k+1|k}$ for an I/O-SHA upon the arrival of the kth input/output pair. For each $l \in L, d \in D$:

$$\pi_0(l|z^0) = l_0(l) \tag{5}$$

$$p_{1|0}(d_1|z^0, l) = \int_D f_l(d_1|d_0', u_0) d_0(d_0') d(d_0') \tag{6}$$

$$p_{k|k}(d|z^k, l_{k-1}) = \frac{h_{l_{k-1}}(y_k|d, u_k) p_{k|k-1}(d|z^{k-1}, l_{k-1})}{\int_D h_{l_{k-1}}(y_k|d_k, u_k) p_{k|k-1}(d_k|z^{k-1}, l_{k-1}) d(d_k)} \tag{7}$$

$$\pi_k(l|z^k) = \sum_{l_{k-1} \in L} \pi_{k-1}(l_{k-1}|z^{k-1}) \times \int_{D(l_{k-1} \to l|u_k)} p_{k|k}(d_k|z^k, l_{k-1}) d(d_k) \tag{8}$$

$$p_{k+1|k}(d|z^k, l_k) = \frac{1}{\pi_k(l_k|z^k)} \sum_{l_{k-1}} \pi_{k-1}(l_{k-1}|z^{k-1}) \times \int_{D(l_{k-1} \to l_k|u_k)} f_{l_{k-1}}(d|d_k, u_k) p_{k|k}(d_k|z^k, l_{k-1}) d(d_k), \tag{9}$$

where $D(l_{k-1} \to l_k|u_k) \subseteq D$ for each l_k and l_{k-1} is defined as $D(l_{k-1} \to l_k|u_k) := \{d_k \in D : \exists e \in E, o_e = l_{k-1}, t_e = l_k, (u_k, d_k) \in \overline{G}_e\}$, i.e., it is the set of data values that enable the edge from l_{k-1} to l_k while the input is u_k.

Now that we have computed the state probability distribution, given the input/output sequence up to a current time k, we can use this to compute the *likelihood of no-fault*, which is the probability of the refinement G^R being outside of the fault-location $L_m = \{F\}$ and is given by:

$$P_N^k := \sum_{l \notin L_m} \pi_k(l|z^k). \tag{10}$$

Note P_N^k can be found by first computing π_k, which in turn is computed by the filter (5)-(9). A detector issues a fault decision "F" whenever this likelihood of no-fault is lower than a threshold, i.e., when $P_N^k \leq \rho$, and remains silent otherwise. The detector $\mathcal{D} : (U \times Y)^{\mathbb{N}} \to \{F, \epsilon\}$ is formally defined as:

$$\forall z^k \in (U \times Y)^{\mathbb{N}}, [\mathcal{D}(z^k) = F] \Leftrightarrow [\exists j \leq k, P_N^j \leq \rho]. \tag{11}$$

Note that once the detector issues F, it issues F for all subsequent steps, i.e., the detector "doesn't change its mind".

V. Case Study: A Room-Heating Problem

In this section we present the results for fault detection computations presented above by applying to a room heating benchmark, which aims to regulate the temperature in a single room with a single heater, and is inspired from [6], [8]. Let the continuous variable x_k present the room temperature at time k, and the binary variable u_k denote the status of the heater, with $u_k = 1$ if the heater is on at time k and 0 otherwise. The room temperature x_k is assumed to evolve according to the linear stochastic difference equation:

$$x_{k+1} = x_k + a(x_a - x_k) + bu_k + v_k,$$

and the requirement and output variables are given by:

$$r_k = \begin{bmatrix} u_k \\ x_k \end{bmatrix},$$
$$y_k = x_k + w_k,$$

where x_a is the (constant) ambient temperature, and the disturbance v_k and the noise w_k are zero mean Gaussian random variables with variances σ_v^2 and σ_w^2, respectively.

For safety purposes, it is required that the room temperature satisfies $x_l \leq x_k \leq x_h$ for all k. It is also required that the room temperature is guaranteed to be higher than x_w in at most 2 steps after the heater is turned on. Note $x_h > x_w > x_l$ are constants, specified by user/designer. Such correctness requirement can be expressed as LTL formula ϕ:

$$\phi = G[\{x_l < r(2) < x_h\} \wedge \{(r(1) = 1) \Rightarrow (r(2) > x_w) \vee X(r(2) > x_w) \vee XX(r(2) > x_w)\}]. \tag{12}$$

It can be verified that the aforementioned system is prediagnosable with respect to ϕ, and the requirement model R is shown in Fig. 2, which has four states and 9 edges, while reaching the state F indicates the violation of formula (12).

The refinement G^R is such that $L = \{l_0, l_1, l_2, F\}$, $U = \{0, 1\}$, $D = X = Y = \mathbb{R}$, $l_0 = \delta(l_0)$, $d_0 = \delta(x_0)$, $L_m = \{F\}$ and the edges are as shown in Fig. 2. For each $l \in L$,

$$f_l(\cdot|d, u) = \mathcal{N}(\cdot|d + a(x_a - d) + bu, \sigma_v^2), \text{ and}$$
$$h_l(\cdot|d, u) = \mathcal{N}(\cdot|d, \sigma_w^2),$$

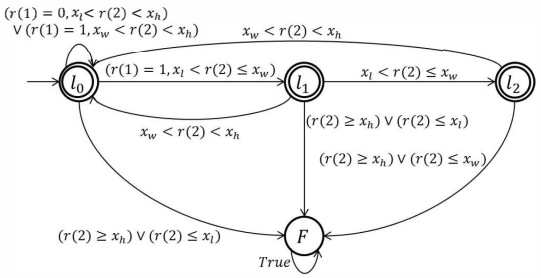

Fig. 2. The requirement model R for single room heating problem.

where $\mathcal{N}(\cdot|\mu, \sigma^2)$ denotes Gaussian distribution with mean μ and variance σ^2. For each $l_j, l_j \in L$ and $u \in U$, $D(l_i \to l_j|u)$ can be easily computed and is shown in Table I.

TABLE I
LIST OF $D(l_i \to l_j|u)$.

$D(l_0 \to l_0\|u=0)$	(x_l, x_h)
$D(l_0 \to l_0\|u=1)$	(x_w, x_h)
$D(l_0 \to l_1\|u=1)$	$(x_l, x_w]$
$D(l_1 \to l_0\|u \in \{0,1\})$	(x_w, x_h)
$D(l_1 \to l_2\|u \in \{0,1\})$	$(x_l, x_w]$
$D(l_2 \to l_0\|u \in \{0,1\})$	(x_w, x_h)
$D(l_0 \to F\|u \in \{0,1\})$	$(-\infty, x_l] \cup [x_h, \infty)$
$D(l_1 \to F\|u \in \{0,1\})$	$(-\infty, x_l] \cup [x_h, \infty)$
$D(l_2 \to F\|u \in \{0,1\})$	$(-\infty, x_w] \cup [x_h, \infty)$
$D(F \to F\|u \in \{0,1\})$	$(-\infty, \infty)$
Others	$\emptyset$

For the computational study, we set $x_a = 70$, $a = 0.1, b = 3, \sigma_v^2 = \sigma_w^2 = 0.4$, and suppose the system is initialized at $x_0 = 80$ and l_0. Suppose the specification parameters are $x_l = 70$, $x_h = 90$ and $x_w = 80$. For simulation, the continuous space is discretized by a grid size of 0.1 over the range $[65, 100]$. The input is such that the heater switches between on and off at each discrete time. A total of 5000 runs, with terminal time $T = 200$, were simulated, out of which there were 457 runs violating the correctness requirement. We implemented the detection algorithm (5)-(11), and the results are shown in Figs. 3-5. In Fig. 3, the room temperature exceeds the upper limit, whereas in Fig. 4, the correctness requirement is violated since the room temperature remains below $x_w = 80$ two steps after the heater is on. In both cases, the likelihood of no-fault, P_N, drops soon after the specification model R reaches state F, and the fault can be detected with a delay of 7 steps by using a detection threshold $\rho < 0.5$. The performance of the detection procedure can be evaluated by the errors in terms of false alarms and missed detections (formally defined in next section), and Fig. 5 shows the number of runs that are false-alarmed or missed-detected over the 5000 runs, as the detection threshold ρ and detection delay n are changed.

VI. PERFORMANCE EVALUATION AND STOCHASTIC DIAGNOSABILITY

Here we formally define false alarm (FA) and missed detection (MD) rates, by first introducing the following

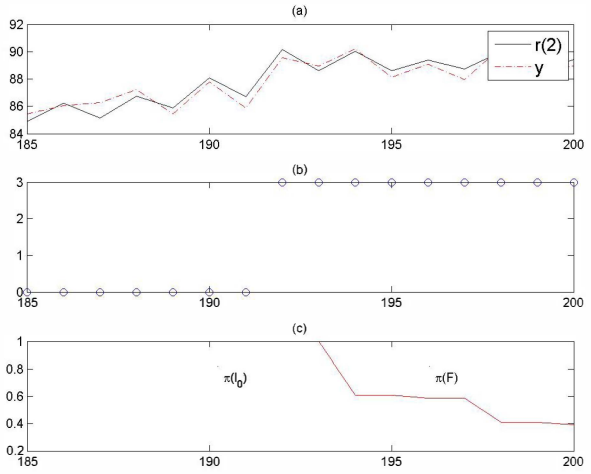

Fig. 3. The detection result for a run that violates the correctness requirement by exceeding the upper limit x_h. (a) true $r(2) = x$ v.s. $y = x + w$; (b) the true state of specification model R where the fault-location F is represented by the number 3; (c) the estimate of state probability distribution.

notions.

A finite run of the system is a finite execution of the stochastic difference equations (1)-(3), denoted as $\overline{z} := (u^{|\overline{z}|}, x^{|\overline{z}|}, r^{|\overline{z}|}, y^{|\overline{z}|})$, where $|\overline{z}| < \infty$ and for each $o \in \{u, x, r, y\}$, $o^{|\overline{z}|} := (o_0, \ldots, o_{|\overline{z}|})$. A run is a fault-run if $r^{|\overline{z}|} \in I_\phi(G)$, where recall that $I_\phi(G)$ is the set of all indicators, and otherwise it is a nonfault-run. Given two runs $\overline{z}_1 := (u_1^{|\overline{z}_1|}, x_1^{|\overline{z}_1|}, r_1^{|\overline{z}_1|}, y_1^{|\overline{z}_1|})$ and $\overline{z}_2 := (u_2^{|\overline{z}_2|}, x_2^{|\overline{z}_2|}, r_2^{|\overline{z}_2|}, y_2^{|\overline{z}_2|})$, $\overline{z}_1$ is said to be a prefix of $\overline{z}_2$, denoted as $\overline{z}_1 \leq \overline{z}_2$, if $|\overline{z}_1| \leq |\overline{z}_2|$ and $o_2^{|\overline{z}_1|} \equiv o_1^{|\overline{z}_1|}$ for each $o \in \{u, x, r, y\}$. In this case we denote $\overline{z}_2 \backslash \overline{z}_1$ as an extension of $\overline{z}_1$. Associated with each run $\overline{z}$ is a sequence of detection statistics, $P_N^0, P_N^1, \ldots, P_N^{|\overline{z}|}$, computed through (5)-(10).

A false alarm (FA) occurs if the detector issues "F" for a nonfault-run, and a missed detection (MD) occurs if the detector remains silent n steps after the system executes an indicator, where n is the detection delay bound allowed by the detector. So the rates of FA and MD can be defined as:

$$P^{fa} := Pr(\overline{z} : r^{|\overline{z}|} \notin I_\phi(G) \wedge P_N^{|\overline{z}|} \leq \rho) \quad (13)$$

$$P^{md} := Pr(\overline{z} : \exists k < |\overline{z}| - n, r^k \in I_\phi(G), P_N^{|\overline{z}|} > \rho) \quad (14)$$

In the following we present a characterization of the class of systems for which detectors with arbitrary accuracies can be designed, by introducing the notion of *Stochastic-Diagnosability* which requires that for any tolerable threshold ρ and error bound τ, there must exist a delay bound n such that for any fault-run, its extensions, longer than n and having likelihood of no-fault lower than ρ, occur with probability at most τ.

Definition 3: Given a system G subjected to an input-sequence drawn from a distribution μ, with correctness requirement expressed in LTL formula ϕ, (G, μ, ϕ) is said to be *Stochastically-Diagnosable*, or simple *S-Diagnosable*,

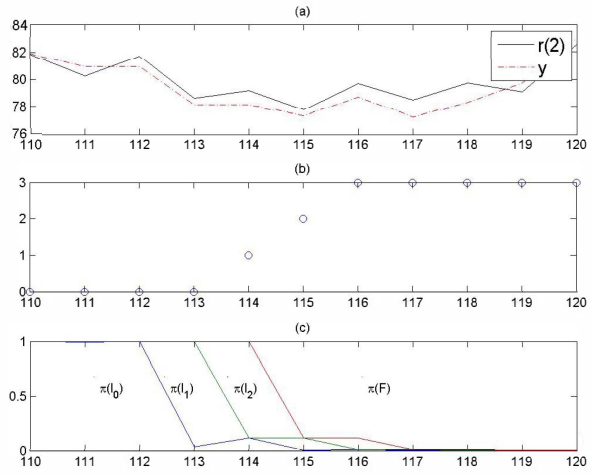

Fig. 4. The detection result for a run that violates the correctness requirement by failing to reach x_w within 2 steps after the heater is on. (a) true $r(2) = x$ v.s. $y = x + w$; (b) the true state of specification model R where the fault-location F is represented by the number 3; (c) the estimate of state probability distribution.

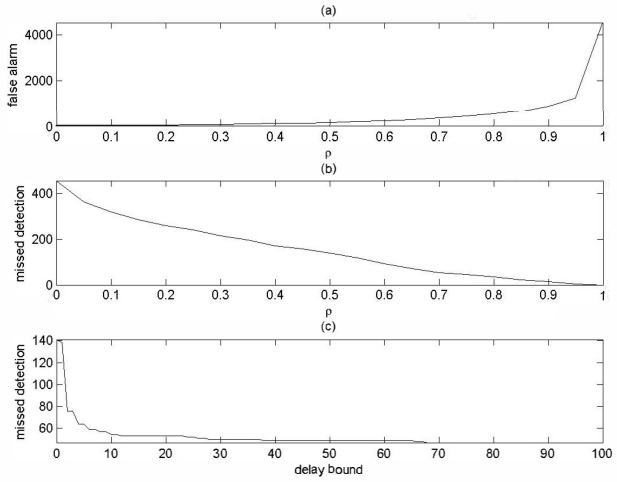

Fig. 5. (a) The number of false alarms as a function of the threshold; (b) the number of missed detections as a function of the threshold; (c) the number of missed detections as a function of detection delay, when the threshold is $\rho = 0.75$.

if $\forall \rho, \tau > 0$, $\exists n \in \mathbb{N}$, such that for any fault-run $\overline{z}_0$, $Pr(\overline{z} \backslash \overline{z}_0 : |\overline{z}| - |\overline{z}_0| > n, P_N^{|\overline{z}|} > \rho) < \tau$.

The following theorem establishes the significance of the S-Diagnosability property, by showing its necessity and sufficiency for the existence of a detector to achieve any desired level of accuracy as measured in terms of FA and MD rates. The proof is omitted here for the sake of space.

Theorem 1: For any FA rate $\nu > 0$ and MD rate $\tau > 0$, there exists a detection threshold ρ and delay bound n so that the rates of FA and MD defined by (13)-(14) satisfy $P^{fa} < \nu$ and $P^{md} < \tau$ if and only if (G, μ, ϕ) is S-Diagnosable.

Remark 5: Theorem 1 identifies the class of systems for which a detector of any desired accuracy can be constructed. Therefore, the S-Diagnosability property should be checked before designing a detector—A desired accuracy may not be achievable if S-Diagnosability is not satisfied. The future work will focus on the verification of S-Diagnosability, together with algorithm that computes a detector so as to ensure the desired rates of FA and MD. ■

VII. Conclusion

This paper studied the fault detection of discrete-time stochastic systems subject to linear-time temporal logic correctness requirement. The continuous physical system (modeled as stochastic difference equations) was refined against its LTL correctness requirement to yield an input-output stochastic hybrid automaton which preserves the behavior of the physical system and captures the requirement-violation as a reachability property. Based on this refinement, the likelihood of no-fault was recursively computed for issuing a detection decision: a fault decision is issued when the likelihood of no-fault drops below a suitably chosen threshold. The performance of the diagnosis procedure was evaluated in terms of false alarm and missed detection rates, and the existence of detector to achieve any desired false alarm and missed detection rates was captured as Stochastic-Diagnosability introduced in this paper. In future, the analytical computation of the rates of false alarm and missed detection will be investigated, together with the verification of the Stochastic-Diagnosability property.

References

[1] D. Thorsley and D. Teneketzis, "Diagnosability of stochastic discrete-event systems," *IEEE Trans. Autom. Control*, vol. 50, no. 4, pp. 476–492, Apr. 2005.

[2] J. Chen and R. Kumar, "Polynomial test for stochastic diagnosability of discrete event systems," *IEEE Trans. Auto. Sci. and Eng.*, vol. 10, no. 4, pp. 969–979, Oct. 2013.

[3] ——, "Online failure diagnosis of stochastic discrete event systems," in *Proc. 2013 IEEE Multi-Conf. Syst. and Control*, Hyderabad, India, Aug. 2013, pp. 194–199.

[4] ——, "Decentralized failure diagnosis of stochastic discrete event systems," in *Proc. 9th IEEE Int. Conf. Autom. Sci. and Eng.*, Madison, WI, Aug. 2013, pp. 1083–1088.

[5] R. H. Chen, D. L. Mingori, and J. L. Speyer, "Optimal stochastic fault detection filter," *Automatica*, vol. 39, no. 3, pp. 377–390, Mar. 2003.

[6] A. Abate, S. Amin, M. Prandini, J. Lygeros, and S. Sastry, "Computational approaches to reachability analysis of stochastic hybrid systems," *Hybrid Systems: Computation and Control*, vol. 4416 of LNCS, pp. 4–17, 2007.

[7] M. Li and R. Kumar, "Reduction of automated test generation for simulink/stateow to reachability and its novel resolution," in *Proc. 9th IEEE Int. Conf. Autom. Sci. and Eng.*, Madison, WI, Aug. 2013.

[8] A. Fehnker and F. Ivančić, "Benchmarks for hybrid systems verification," *Hybrid Systems: Computation and Control*, vol. 2293 of LNCS, pp. 326–341, 2004.

[9] C. Baier and J.-P. Katoen, *Principles of model checking*. MIT Press, 2008.

[10] S. Jiang and R. Kumar, "Failure diagnosis of discrete-event systems with linear-time temporal logic specifications," *IEEE Trans. Autom. Control*, vol. 49, no. 6, pp. 934–945, Jun. 2004.

[11] ——, "Diagnosis of repeated failures for discrete event systems with linear-time temporal-logic specifications," *IEEE Trans. Auto. Sci. Eng.*, vol. 3, no. 1, pp. 47–59, Jan. 2006.

[12] S. Summers and J. Lygeros, "Verification of discrete time stochastic hybrid systems: A stochastic reach-avoid decision problem," *Automatica*, vol. 46, no. 12, pp. 1951–1961, Sep. 2010.